Response to call for input for EMRTD Study "Artificial Intelligence, Cultural Rights, and the Right to Development"

Issued by Expert Mechanism on the Right to Development United Nations Human Rights Office of the High Commissioner (OHCHR)

29 November 2025

About this Response

We are submitting this response to the Expert Mechanism on the Right to Development (EMRTD) United Nations Human Rights Office of the High Commissioner (OHCHR) inquiry about Artificial Intelligence, Cultural Rights, and the Right to Development, on behalf of the CulturAI project. We are a team of researchers funded by Responsible AI UK (RAI UK)[1], working on Human-Computer Interaction (HCI) and Responsible Artificial Intelligence (AI) topics, and based at the University of Nottingham, University of Southampton, and King's College London. CulturAI is a 12-month project, running from February 2025 to February 2026. We have been investigating the evaluation of cultural sensitivity and representation in Text-to-Image (T2I) models. We welcome this consultation and aim to inform the EMRTD study.

The response is based on our research findings[2] from a literature review on how culture has been conceptualised in Generative AI (Gen AI) research, and co-creation workshops with 59 individuals from diverse cultural backgrounds in 19 countries, in which we co-created a mixed-methods methodology for evaluating cultural representation in T2I models. This response also draws on our own experiences and expertise.

Suggested Citation

Reyez-Cruz, G., Kiden, S., Azim, T., Bergin, A. G., Choi, S., Eke, D., Klyshbekova, M., Peter, O., Waheed, M., Devlin, K., Fischer, J. E., Ramchurn, S., Stein, S., Vallejos, E. P. (2025). Responsible AI UK's Response to Call for Input for EMRTD Study "Artificial Intelligence, Cultural Rights, and the Right to Development". Responsible AI UK. DOI: 10.25878/7nmj-0t79

Executive Summary

In our response below, we have combined some questions and addressed them collectively, covering the following topics:

Benefits, existing digital divides, and risks of AI for cultural rights, cultural participation and cultural diversity

- Generative AI (GenAI) systems, including Text-to-image (T2I) models, such as Midjourney, Dall-E, Stable Diffusion and Gemini, could enhance cultural participation and diversity by empowering

[1] RAI UK https://rai.ac.uk/

[2] Kiden, S., Peter, O., Reyes-Cruz, G., Klyshbekova, M., Choi, S., Bergin, A. G., ... & Fischer, J. E. (2025). Back to the Communities: A Mixed-Methods and Community-Driven Evaluation of Cultural Sensitivity in Text-to-Image Models. https://arxiv.org/abs/2510.27361

communities to represent and promote their cultures. They could also help to address cultural underrepresentation in digital and physical media **[par. 1-2]**.

- However, current T2I models default to Western imagery and often produce stereotyped outputs, for instance by perpetuating diminishing visual narratives about developing and least developed countries **[par. 3-6].**
- Already identified harms by Artificial Intelligence (AI) include representational harms, allocative harms, quality-of-service harms, interpersonal harms and social harms. Other harms in the context of T2I models include exoticism and cultural misappropriation **[par. 7-9].**
- Under-represented groups, including communities in the Global South/Majority and indigenous and minority groups in the Global North, are disproportionately affected by harms emerging from T2I models such as misrepresentation, exoticism, cultural misappropriation and quality-of-service harms **[par. 10-11]**.

Protection of cultural rights in the development of AI systems

- Long-term effects of T2I model use on cultural rights include visual cultural loss, flattening of cultural richness and weakening cultural self-determination **[par. 12-14]**.
- Strategies for cultural rights protection should involve co-creation and co-production with affected cultural communities across the data collection, testing, and system design and evaluation stages **[par. 15-16]**.
- Embedding Responsible Research and Innovation (RRI) principles and strengthening Equality, Diversity, and Inclusion (EDI) processes can support the goal of cultural preservation in the era of rapid AI development, for instance, by implementing cultural impact assessments throughout the AI lifecycle **[par. 17-19]**.

Recommendations for the regulation of AI

- Regulations can provide cultural safeguards and meaningful participation from affected communities. Incentives are needed for AI companies to be more mindful of cultural rights **[par. 20-21]**.
- "Low-risk" or consumer-grade tools should also be considered for regulation, and this should not be purely technical or risk-based. For instance, decolonial impact assessments could be adopted to address these aims. **[par. 22-24]**.
- Self-regulation by technology companies is not sufficient to protect cultural rights when pursuing the right to development **[par. 25-28]**.
- Regional agreements are not yet equipped to address complex and rapidly evolving cultural risks introduced by AI **[par. 29]**.

Key Questions

Question 1. In your opinion, what, if any, are the potential benefits of Artificial Intelligence (AI) for cultural rights in the context of the right to development? You may, for instance, consider AI's impact on development, cultural participation, cultural diversity (including language preservation, artistic creation and expression, access and participation in science, academic and scientific freedom, and the protection of moral and material rights of authors and creators)?

1. Artificial Intelligence (AI) technologies, including GenAI have the potential to promote cultural participation and cultural diversity. Among these are Text-to-Image (T2I) models such as Midjourney,

Dall-E, Stable Diffusion and Gemini, which receive input in the form of text and display outputs in the form of images or illustrations. T2I models have have become widely accessible[3] to various populations across the world, providing them with the means to easily generate digital media at a low or no cost. Whilst most T2I models providers do not publish usage statistics, it has been estimated that around 34 million images[4] are being generated every day, used for a range of purposes such as brainstorming creative ideas for art, logos, and character designs to developing marketing materials and creating illustrations for education and research.

2. The potential benefits of these tools include supporting communities in representing themselves, promoting their cultures, and working towards the reduction of the lack of diverse cultural representation in digital and physical media. Nonetheless, current research indicates that T2I models consistently produce biased outputs. Such biases have been largely documented, including gender, skin tone, and other racial biases[5,6]. This has led to an effort to understand and evaluate the sensitivity of cultural representation in T2I models, as well as the harms they can generate, so to raise awareness amongst different stakeholders (e.g. end-users, developers, designers, policymakers) and actions to address these existing shortcomings (e.g. user guidelines and technical improvements).

Question 3. To what extent, if any, do existing digital divides deprive developing and least developed countries from reaping those benefits? Question 4. Can you provide any specific real-life examples involving the impacts of such digital divides on the enjoyment of cultural rights when pursuing the right to development?

3. T2I models are primarily trained on large datasets scraped from the Internet, and therefore reflect dominant (often Western) visual norms. That is, using neutral prompts that do not specify a specific culture will often show Western imagery. For instance, prompting for an image of a standard breakfast will likely generate a plate of eggs, bacon and pancakes in an American style. When communities from the Global South, and indigenous or underserved groups in the Global North attempt to use these tools to depict their traditions and everyday activities, outputs will either a) default to Western styles, b) produce stereotyped images, or c) fail entirely, undermining communities' capacity to represent and promote their culture in digital media.[7,8,9]

4. Findings from our research[10] indicate that current T2I models replicate existing visual narratives about developing and least developed countries. An example comes from our workshop sessions with individuals from diverse cultural backgrounds and expertise, in which a T2I model was used to generate images with prompts they suggested. When prompting the model to depict images of *"children engaging with technology"* initially in a neutral or no specific context and subsequently

[3] Steinfeld, K. (2023). Clever little tricks: A socio-technical history of text-to-image generative models. International Journal of Architectural Computing. https://doi.org/10.1177/14780771231168230
[4] Valyaeva, A. (2023). AI Image Statistics for 2024: How Much Content Was Created by AI. https://journal.everypixel.com/ai-image-statistics
[5] Cho, J., Zala, A., & Bansal, M. (2023). DALL-EVAL: Probing the Reasoning Skills and Social Biases of Text-to-Image Generation Models. IEEE/CVF International Conference on Computer Vision (ICCV). https://doi.org/10.1109/ICCV51070.2023.00283
[6] Fraser, K., & Kiritchenko, S. (2024) Examining Gender and Racial Bias in Large Vision–Language Models Using a Novel Dataset of Parallel Images. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.eacl-long.41
[7] Ghosh, S., & Caliskan, A. (2023). 'Person' == Light-skinned, Western Man, and Sexualization of Women of Color: Stereotypes in Stable Diffusion. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.465
[8] Ghosh, S., Lutz, N., & Caliskan, A. (2024). "I Don't See Myself Represented Here at All": User Experiences of Stable Diffusion Outputs Containing Representational Harms across Gender Identities and Nationalities. https://doi.org/10.1609/aies.v7i1.31650
[9] Ghosh, S., Gautam, S., Venkit, P. N., & Ghosh, A. (2025). Documenting Patterns of Exoticism of Marginalized Populations Within Text-to-Image Generators. https://doi.org/10.1609/aies.v8i2.36614
[10] See Ref 2.

in specific parts of the world or about specific cultures, participants noted that children in Sub-Saharan Africa were often depicted in rural settings, which heavily rely on common conceptions about the Global South. Cultural experts in the team suggest that data used to trained T2I models may have included imagery distributed by humanitarian organisations, which could influence the portrayal of charity-related narratives in the AI-generated images. While this type of visual representation somewhat reflects the daily socio-economic realities in rural Africa, workshop participants also expressed concerns that this *"regimented"* depiction may have negative implications for the communities, ultimately limiting how these communities are perceived.

5. Our work has also shown that stereotypes emerge in AI-generated images across the globe, including cultures in both the Global North and Global South. However, we also established that not all cultural stereotypes are considered demeaning by their members. For instance, whilst Swiss participants recognised that stereotypes depicted in the AI-generated images about Swiss culture are often used by the government for promoting tourism (e.g. mountains and natural landscapes, clothing, cheese), Ugandan participants raised concerns about the consistent way their culture is portrayed in the AI-generated images (e.g. extremely rural landscapes even when prompted for urban settings) which may lead outsiders to develop misconceptions (e.g. that there are no urban spaces whatsoever in the country) and may discourage potential visitors from considering the region as a travel destination. In another example, participants from Ireland noted that images generated, including those of a family, typically included what appeared to be a drink of Guinness; whilst at first participants found this amusing, the sheer number of images of Irish culture containing alcoholic beverages perpetuates unhealthy stereotypes.

6. Considering the use of AI to produce images for use in online and offline contexts such as social media or advertising, and how quickly these can be produced, the proliferation of misrepresentations and potentially negative or harmful images impacts on these communities disproportionately as it leads to the perpetuation of stereotypes.

Question 5. What are the main risks posed by and drawbacks already identified of Artificial Intelligence, including, amongst others, generative AI, to cultural rights in pursuing the right to development? Question 6. In addition to the above, please set out your views on the following potential AI risks and drawbacks in terms of how they relate to cultural rights: (i) Algorithmic bias.

7. Past work has remarked on the importance of explicitly identifying, reporting, and labelling the specific types of harms that AI research aims to address. Discussions surrounding algorithmic harms and their perpetuation are interconnected[11]. Clearly defining these harms can serve as an effective tool for systematically identifying, analysing, and mitigating them, whereas broadly categorising them as *'algorithmic bias'* risks obscuring the full scope of their impacts and the groups most affected. Shelby et al.[12] propose categorising algorithmic harms into five primary types:

    - *Representational harms* relate to how algorithmic systems depict, or fail to depict, certain communities. Examples include reinforcing stereotypes, generating demeaning or alienating imagery, or completely excluding these groups from system outputs.
    - *Allocative harms* occur when these systems influence the distribution of resources, such as access to social services or employment opportunities, systematically excluding or disadvan-

[11] Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (Technology) is Power: A Critical Survey of "Bias" in NLP. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.485

[12] Shelby, R., et al. (2023). Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. https://doi.org/10.1145/3600211.3604673

taging specific communities, thereby limiting economic or social opportunities.

- *Quality-of-service harms* happen when a system is optimised for one group but performs inadequately for others. This may result in feelings of alienation or require additional effort from affected users to compensate for system deficiencies.
- *Interpersonal harms* involve the way technology mediates relationships between individuals or groups. Examples include technology-facilitated stalking, surveillance, domestic abuse, or increased online harassment.
- *Social harms* encompass the wider societal impacts of these systems, such as algorithm-driven political polarisation, misinformation dissemination, or environmental effects stemming from large-scale computing infrastructure.

8. In the context of T2I models, Ghosh et al.[13] have defined specific types of representational harms related to culture:

    - *Exoticism* refers to the overemphasis or excessive representation of certain features or characteristics in general depictions of a culture, often at the expense of maintaining culturally accurate details.
    - *Cultural misappropriation* refers to inaccurate blending of elements from different cultures or sub-cultures.

9. Our study highlighted the discomfort that participants felt when aspects of their culture, of their ethnicity, heritage, countries, etc., were misrepresented or negatively portrayed. Much like we are only now grasping the significant impact that social media has on mental health, particularly for young women, the potential representational harms from T2I models are likely to be felt most by future generations. Parallels can be drawn with the lack of diversity in media and the responses to attempts to tackle this.

Question 7. Do those risks and drawbacks disproportionately affect any particular category of individuals or groups of people when pursuing their right to development? Please explain below.

10. Yes, as highlighted above, under-represented groups, including communities in the Global South, and indigenous and minority groups in the Global North are disproportionately affected by harms emerging from T2I models[14]. In our co-creation workshops, sometimes image outputs were either exoticised or inaccurate depictions about specific groups. These included imagery about the Rohingya community, or portrayals of Somali food which was blended into neighbouring Ethiopian culture. It is likely to be felt by members of smaller communities, not just specific countries or ethnicities but including other cultural groups. Although our research has so far focused on representation from different countries, findings from smaller countries like Ireland and Switzerland, e.g., that T2I models produce Irish-themed birthdays when prompted for an Irish birthday, indicate that there may be some obfuscation of smaller communities in preference of majority views. It is also concerning when considering what majority views may be represented in the training data of these types of models.

11. Moreover, as T2I models produce Western imagery by default, users who are not part of WEIRD (Western, Educated, Industrialised, Rich, Democratic) populations[15] require multiple refinements

[13] Ghosh, S., Venkit, P. N., Gautam, S., Wilson, S., & Caliskan, A. (2024). Do Generative AI Models Output Harm while Representing Non-Western Cultures: Evidence from A Community-Centered Approach. https://doi.org/10.1609/aies.v7i1.31651
[14] See Ref 9.
[15] Mihalcea, R., et al. (2025). Why AI Is WEIRD and Shouldn't Be This Way: Towards AI for Everyone, with Everyone, by Everyone. https://doi.org/10.1609/aaai.v39i27.35092

and iterations of text prompts to generate appropriate image outputs. This process (of several iterations) places an unfair burden on these users, not only incurring additional computational costs, but also undermines and alienates the users, who experience quality-of-service harms. These groups face higher risks of misrepresentation, exclusion, and bias in data-driven systems. Incomplete datasets or biased images can lead automated decision-making tools to worsen these problems. For instance, prior research[16,17] shows that facial recognition systems often vary in performance across gender and racial lines, usually being more accurate for white males than women of colour or other minorities.

Question 8. What do you believe might be the long-term effects of AI use on cultural rights and, in that context, the future of the right to development, including cultural self-determination?

12. T2I models trained primarily on online datasets that are not curated, which are unlikely to represent the plurality and diversity of cultures and can contain inaccurate visuals or be primarily selected or produced by people not representing that culture, risk becoming the dominant digital source of *"what a culture looks like"*. These skewed cultural depictions may circulate widely in digital and physical spaces, (e.g. archives, search engines, educational materials, media), be used further in training AI models, and can accelerate cultural loss, displacing local knowledge and authentic community-generated imagery. This dynamic undermines cultural preservation and restricts the ability of communities to define their cultural identities on their own terms or have multiple versions of that culture represented.

13. Obfuscation of smaller cultural groups, of the different ways culture can be represented even within the same group, in T2I models can lead to the flattening of culture by hiding the variety and diversity that bring richness and value. Only those with experience of those cultures will be able to identify its loss within images but are unlikely to be able to stem the tide once AI-generated, bland images proliferate.

14. Additionally, these models also risk weakening cultural self-determination by allowing third parties to generate and commercialise imagery derived from cultural materials without consent, while normalising inaccurate portrayals as authentic. As AI-generated depictions increasingly shape public understanding of cultural groups, communities may find themselves struggling to correct misconceptions or maintain control over their own cultural narratives. This can reinforce global power imbalances by privileging dominant cultural aesthetics while erasing or homogenising minority ones.

Question 9. How can cultural rights be protected in the era of rapid AI development ? You may, for example, consider prevention and mitigation.

15. Protecting cultural rights in the era of rapid AI development requires multiple measures that support prevention, early identification of risks and harms and ongoing measures of mitigation. While not exhaustive, several actions are particularly relevant such as the use of co-creation and co-production approaches with affected cultural communities as also done in our study[18]. This entails involving them directly in decisions about data collection, testing, and system design and evaluation, acknowledging diversity within cultural groups, and requires community members to be treated as partners rather than subjects to be studied. It can reduce misrepresentation, likeli-

[16] Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Machine Learning Research https://proceedings.mlr.press/v81/buolamwini18a.html
[17] Ferrante, E. (2021). Inteligencia artificial y sesgos algorítmicos¿ Por qué deberían importarnos?
[18] See Ref 2.

hood of cultural appropriation, and provide proper channels for communities to correct inaccuracies through iterative feedback rather than simply one-off consultations. It also promotes transparency about how community input informs design choices and helps extend participation to those who may otherwise be excluded. Put simply, *"nothing about us, without us"*[19].

16. It is imperative that when considering the value of these co-production approaches, that all steps of the AI lifecycle are considered. Our role as researchers limits our capacity to bring meaningful change at the level of model development without access to them, but we are still able to evaluate deployed models and highlight where change can happen. Technology companies should listen and continue model development responding to issues raised.

17. Embedding of Responsible Research and Innovation (RRI) principles guided by the UK Anticipate, Reflect, Engage, Act (AREA) framework[20] can promote transparency while fostering early anticipation of cultural risks, supporting ongoing reflection and meaningful engagement with communities. When applied thoroughly, RRI can be a helpful lens for identifying cultural nuances and highlighting blind spots in data or system behaviour and help prevent representational harms before they occur. It can also help redistribute influence by identifying and involving under-represented or even overlooked groups and ensure that AI development upholds cultural diversity rather than reinforcing existing imbalances.

18. Strengthening Equality, Diversity, Inclusive (EDI) processes is also essential for protecting cultural rights and requires diversifying AI teams to reduce over-reliance on only certain cultural perspectives and removing systemic barriers that prevent meaningful participation from under-represented groups[21]. Conducting internal checks for bias and misrepresentation across data collection and evaluation processes is needed alongside recognizing that community based knowledge can serve as a form of expertise that can help inform system design positively. Stronger EDI processes can lead to organisations providing the right support, policies and resources that ensure cultural rights are considered throughout development rather than as an afterthought.

19. Lastly, cultural impact assessments should also be a part of AI development. These assessments can evaluate how AI systems may represent, misrepresent or even omit cultural groups and potential harms that may rise[22,23]. The need is for analysing training data to uncover gaps or imbalances in how different cultures are represented, and incorporate systematic evaluation of model outputs across diverse cultural scenarios to understand how systems behave in practice. Assessments should be done for the full lifecycle of AI models, meaning a) before deployment to anticipate risks and harms, b) after deployment to monitor real world performance and impact and c) through ongoing monitoring as models evolve and new unseen challenges emerge. This further needs to include mechanisms to collect, receive and respond to feedback from diverse cultural groups. Furthermore, proper mitigation measures need to be established such as revising datasets or refining the model behaviour and or including cultural guardrails, while documenting and sharing the decision process openly and transparently.

[19] https://en.wikipedia.org/wiki/Nothing_about_us_without_us

[20] UK Research and Innovation. (2023).https://www.ukri.org/who-we-are/epsrc/our-policies-and-standards/framework-for-responsible-innovation/

[21] Gilman, M. E. (2023). Policy Brief on Democratizing AI: Principles for Meaningful Public Participation. https://papers.ssrn.com/abstract=4628755

[22] Eke, D., Chavarriaga, R., & Stahl, B. (2025). Decoloniality impact assessment for AI. https://link.springer.com/article/10.1007/s00146-025-02649-4.

[23] Hendriks, P., Sturm, T., Mehler, M. F., & Buxmann, P. (2024). The Impact of Artificial Intelligence on the Evolution of Culture. https://www.researchgate.net/publication/385379944_The_Impact_of_Artificial_Intelligence_on_the_Evolution_of_Culture

Question 10. Do you think regulating AI would be an effective way to protect cultural rights when pursuing the right to development?

20. The right to development is fundamentally about enabling individuals and communities to participate in, contribute to, and benefit from economic, social, cultural and political development. Cultural rights are central to ensuring that this is inclusive, equitable and respectful of diverse ways of life. Ethical principles and soft laws that are optional are not enough to guarantee these rights[24]. However, regulations can help to move cultural rights from aspirational statements into legal obligations. This is particularly relevant for societies in the Global Majority where AI is eroding cultural rights. AI systems often operate on assumptions, values and datasets that reflect dominant cultural, linguistic and epistemic norms which can unintentionally underline cultural rights as underscored in Article 15 of the International Covenant on Economic, Social and Cultural Rights (ICESCR)[25], the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP)[26], and the United Nations Educational, Scientific and Cultural Organisation (UNESCO) Universal Declaration on Cultural Diversity[27]. There is sufficient literature on elements of this cultural erosion from linguistic marginalisation[28], data and algorithmic colonialism[29]. Regulations can provide cultural safeguards, or meaningful participation from affected communities.

21. It is important that we also explore how to incentivise AI companies to be more mindful of cultural rights. Organisationally, the easy path is the one we find ourselves on. It is imperative that rather than relying on reactive solutions or ”the stick” of enforcement, that we also consider how to make use of incentives, the ”carrot”. One way is through reputational strategies. For example, several social media platforms have worked with charities such as Samaritans to tackle online harms, providing funding and support with dissemination. Likewise, regulation can be a barrier to competition, especially where smaller companies are unable to compete with high cost or high risk development and deployment cycles. Providing standards or guidance, frameworks that can be followed, are an important additional requirement to regulatory processes. Within our project, CulturAI, we are aiming to develop methodologies for participatory evaluation.

Question 11. If so, what kinds of AI uses or tools should be regulated, how, and by whom?

22. Given AI’s pervasive influence across economic, cultural, political, and epistemic domains, all AI systems require regulation. This necessity extends beyond high-risk applications to include seemingly “low-risk” or consumer-grade tools, as their cumulative, systemic, and cultural impacts, particularly where historical inequalities exist, demand oversight.

23. Regulation must not be purely technical or risk-based. It must be guided by a value-based strategic governance paradigm that aligns AI with societal values, cultural rights, and the right to development. A value-based paradigm embeds shared societal values into the full lifecycle of AI, including: human dignity, cultural rights and diversity, equity and justice, sustainability and intergenerational justice, democratic participation, epistemic pluralism (recognising multiple knowledge systems).

[24] Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

[25] ICESCR https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-economic-social-and-cultural-rights

[26] UNDRIP https://www.ohchr.org/en/indigenous-peoples/un-declaration-rights-indigenous-peoples

[27] UNESCO cultural diversity norms https://www.unesco.org/en/legal-affairs/unesco-universal-declaration-cultural-diversity

[28] Helm, P., Bella, G., Koch, G., & Giunchiglia, F. (2024). Diversity and language technology: How language modeliing bias causes epistemic injustice. Ethics and Information Technology, 26(1), 8. https://doi.org/10.1007/s10676-023-09742-6

[29] Birhane, A. (2020). Algorithmic Colonization of Africa. SCRIPTed: A Journal of Law, Technology & Society, 17(2). https://doi.org/10.2966/scrip.170220.389

24. A decoloniality impact assessment should be required for all AI development and deployments in the Global Majority[30]. This involves assessing:

    - Epistemic impacts: Does the AI system privilege certain knowledge systems? Does it marginalise indigenous or local knowledge?
    - Cultural impacts: Does it appropriate or distort cultural artefacts? Does it contribute to linguistic or cultural erosion?
    - Power and governance impacts: Who benefits and who is burdened? Does it reinforce colonial power asymmetries?
    - Data sovereignty and ownership: Do communities control their cultural and personal data?
    - Representation and participation: Are historically marginalised groups part of design and oversight?

Question 12. Is self-regulation of technology companies that develop AI sufficient to protect cultural rights? If not, why not?

25. No. Self-regulation by technology companies is not enough to safeguard cultural rights. Empirical evidence, from social media platforms to AI systems, reveals persistent harms,[31] especially among young people, due to persuasive design and addictive features. Legislative frameworks like the UK's Online Safety Act[32] have yet to be enforced effectively, leaving cultural and societal protections unaddressed.

26. The UN Guiding Principles on Business and Human Rights[33] establish a corporate responsibility to respect human rights through due diligence processes, including assessment, mitigation, and remediation of harms caused by business operations—AI included. However, reliance on voluntary self-assessment allows commercial incentives to overshadow cultural rights, and prioritising actions that are convenient rather than those that genuinely benefit society. External and binding oversight is needed to ensure genuine accountability.

27. The UNESCO Recommendation on the Ethics of AI[34] affirms that AI must respect human rights, including cultural diversity and freedom of expression. It outlines actionable policy areas, such as cultural governance and embracing pluralism, yet lacks enforceable mechanisms. Without legally binding obligations, companies are not compelled to prioritise cultural protections.

28. The Human Rights Council has emphasised that both states and companies must adhere to international human rights standards when developing or deploying AI, warning against harms like privacy infringements, discrimination, and cultural erasure. These calls reinforce that self-regulation is insufficient—only externally enforced standards can ensure protection of cultural rights.

Question 15. Are regional agreements equipped to deal with the new AI challenges to cultural rights that underpin or are motivated by the right to development? If not, in what ways do they fall short?

29. No. Whilst many regional human rights and development agreements acknowledge cultural rights,

[30] Eke, D., Chavarriaga, R., & Stahl, B. (2025). Decoloniality impact assessment for AI. AI & SOCIETY. https://doi.org/10.1007/s00146-025-02649-4

[31] Verma, K., Davis, B., Milosevic, T., & Umbach, R. (2025). From Users to Co-Designers: Youth Participation in Understanding Cyberbullying. https://doi.org/10.1145/3713043.3731485

[32] GOV.UK. (24 July 2025). Online Safety Act 2023. https://www.gov.uk/government/collections/online-safety-act

[33] Guiding Principles on Business and Human Rights https://www.ohchr.org/en/publications/reference-publications/guiding-principles-business-and-human-rights

[34] UNESCO Recommendation on the Ethics of AI https://unesdoc.unesco.org/ark:/48223/pf0000381137

they are not yet equipped to address complex and rapidly evolving cultural risks introduced by AI. This is because most of these agreements were drafted before AI became integral to social, cultural and economic life. From the perspective of the Global Majority, these agreements do not address the structural, colonial and developmental dimensions of AI harms. This is a fundamental gap because AI systems amplify epistemic injustice, centralise power in the Global North, especially in the hands of the big tech corporations, extract data from the Global Majority without fair benefits and also reinforce unequal development pathways. Without recognising these structural issues, regional agreements cannot protect cultural rights connected to the right to development.

Response Contributors

Lead Authors (in alphabetical order by first name)

- Dr Aislinn Gomez Bergin, Transitional Assistant Professor, University of Nottingham
- Dr Damian Eke, Transitional Assistant Professor, University of Nottingham
- Dr Gisela Reyez-Cruz, Transitional Assistant Professor, University of Nottingham (Co-Lead)
- Dr Maira Klyshbekova, Postdoctoral Research Associate in AI Education, King's College London
- Dr Maria Waheed, Research Fellow, University of Nottingham
- Oriane Peter, PhD Researcher, King's College London
- Dr Sarah Kiden, Research Fellow, University of Southampton (Co-Lead)
- Sena Choi, Research Assistant, University of Nottingham
- Dr Tayyaba Azim, Visiting Researcher, University of Southampton; AI Evaluation Scientist, University of Cambridge

Advisory Team

- Prof Elvira Perez Vallejos, Professor of Mental Health and Digital Technology, University of Nottingham
- Prof Joel Fischer, Professor of Human-Computer Interaction, University of Nottingham
- Prof Kate Devlin, Professor in AI and Society, King's College London
- Prof Sarvapali (Gopal) Ramchurn, Professor of Artificial Intelligence, University of Southampton
- Prof Sebastian Stein, Professor of Artificial Intelligence, University of Southampton

Please send any queries and comments about this response to gisela.reyescruz@nottingham.ac.uk and info@rai.ac.uk.

About Responsible AI UK

Responsible AI UK brings together researchers from across the four nations of the UK to understand how we should shape the development of AI to benefit people, communities and society. It is an open, multidisciplinary network, drawing on a wide range of academic disciplines. This stems from our conviction that developing responsible AI will require as much focus on the human, and human societies, as it does on AI. Funded by the Technology Missions Fund, we convene researchers, industry professionals, policymakers, and civil society organisations.

About the CulturAI Project

The “CulturAI: Informing Cultural Alignment for Text to Image Generators” project[35] is a Responsible AI UK funded research project, as part of the Cornerstone programme of research. In this project, we investigate the evaluation of cultural sensitivity and representation in Text-to-Image (T2I) models.

[35] CulturAI https://sites.google.com/view/culturaiproject